BAYESIAN STOCKPILE RELIABILITY METHODOLOGY FOR COMPLEX SYSTEMS WITH APPLICATION TO A MUNITIONS STOCKPILE

C. M. ANDERSON-COOK, T. GRAVES, M. HAMADA, N. HENGARTNER,
V. JOHNSON, C. S. REESE, A. G. WILSON

1. Introduction: A Bayesian Approach to Stockpile Reliability

Logistics require the U. S. Armed Forces to purchase weapons, ammunition, and spare parts in large lots that are stored until the time they are used or retired. The management and maintenance of such stockpiles face many issues related to the aging of the stored units. It is well known that, over time, the units degrade to a point that they may fail to function as intended. In the case of a missile system, functioning as intended typically implies that it fires and hits the intended target. The central problem of *stockpile reliability* is to estimate what fraction of units within the stockpile will work. In other words, we are interested in determining the probability that a *randomly* selected unit from the stockpile will function.

Current strategies for assessing the reliability of a stockpile rely on testing the success rate of a sample of units from the stockpile. For expensive and complex systems, the number of sampled units for these destructive full-system tests, which are considered the gold-standard measurements of the success of the system, may necessarily be small. But many other types of tests, from component tests to visual inspection to functional tests of specific subsystems, are possible and can provide indirect information that can complement the full-system testing. Typically, the relative proportion of data available from these alternate sources may dramatically outnumber the full-system tests, and hence their inclusion into analysis methods can enhance prediction and precision of estimation substantially. The challenge is to develop a methodology that allows for the integration of various sources of data and information to assess the performance and reliability of the units in the stockpile.

Los Alamos National Laboratory (LANL) faces related problems in its assessment of the reliability of the stockpile of nuclear weapons. In the absence of full-system nuclear tests, we use data from component and subsystem-level tests, plus computer models and simulations, and integrate that information to produce performance estimates, with associated uncertainty. The integration of component and subsystem data with expert knowledge is essential for obtaining meaningful estimates of reliability upon which policy decisions can be made.

Ideally, we would like to have enough full-system tests to be able to accurately assess the reliability of a stockpile. But due to cost and other considerations, this may not always be possible. Fortunately, there are often other sources of information available about system reliability. For example, complex systems can often be described in terms of interrelated subsystems and components, and test data from component and subsystem tests may be available. Here not only is information about the reliability of pieces of the full-system important, but also the specific nature of how these pieces are connected to each other to produce functionality in the system. (The distinction between a subsystem and a component may not always be obvious. For this paper, components are the smallest subunits we want to test and subsystems are portions of the entire system which represent collections of related components.) Other sources of information also include engineering knowledge related to the components, the subsystems and/or the full system, visual inspections and other non-destructive testing, such as electronic diagnostics and battery checks.

The Statistical Sciences Group at LANL continues to develop a suite of methods and tools, collectively known as Information Integration Technology (IIT), that allow the combination of information from various sources to predict performance and reliability for complex systems with limited or no data from full-system tests. In this paper, we describe one application of IIT to the assessment of performance, reliability and shelf life of units within a stockpile that combines full-system tests with component tests and engineering judgment. The key aspects of this approach that will be explained in some detail in this paper are (1) a unified method for consistently predicting down from the system level that combines

with prediction up from the component level data, and (2) how expert judgement and other indirect sources of knowledge can be incorporated into a single model through informative Bayeisan priors to maximize the information summarized in the overall model.

2. Application: Missile Overview

2.1. **Background and System Model.** The methodology presented has been applied to a particular missile, but for clarity of presentation, we have simplified some details of the system. We precisely describe the modeling and statistical procedures followed, and preserve all of the features of the system that provided statistical challenges for the analysis.

Some of the key characteristics of the system presented include the following. The system can be logically decomposed into subsystems, which are natural groupings of components focused on a particular task of the mission. Some of the subsystems combine directly to form the system, while other subsystems may be nested within subsystems. The form of the system representation is largely driven by where sources of data are available and how the engineers best feel the system functions can be summarized. Data are available at some, but not all, of the components and subsystem. There are also several variants of the missile. Treating each variant as possibly having different reliability can increase the precision of the estimation process and help provide insights into which variants are more likely to fail.

For the example presented in this paper, we consider a system with three major subsystems with a total of 13 components. Figure 1 shows an event tree diagram of how subsystems and components combine to form the system. For more information about event trees, see Roberts (1987). The components and subsystems are connected by AND gates, implying that all components and subsystems need to function for the full system to function. The system is labeled as C_1, comprised of a subsystem C_2 (made up of two components, C_5 and C_6, and a subsystem C_7 with 8 components, C_{10} to C_{17}), a component C_3, and a subsystem C_4 (made up of 2 components C_8 and C_9). As mentioned previously, the labeling of individual parts of the system as components or subsystems is somewhat subjective and is dependent on what information is available. The key factor in determining what level

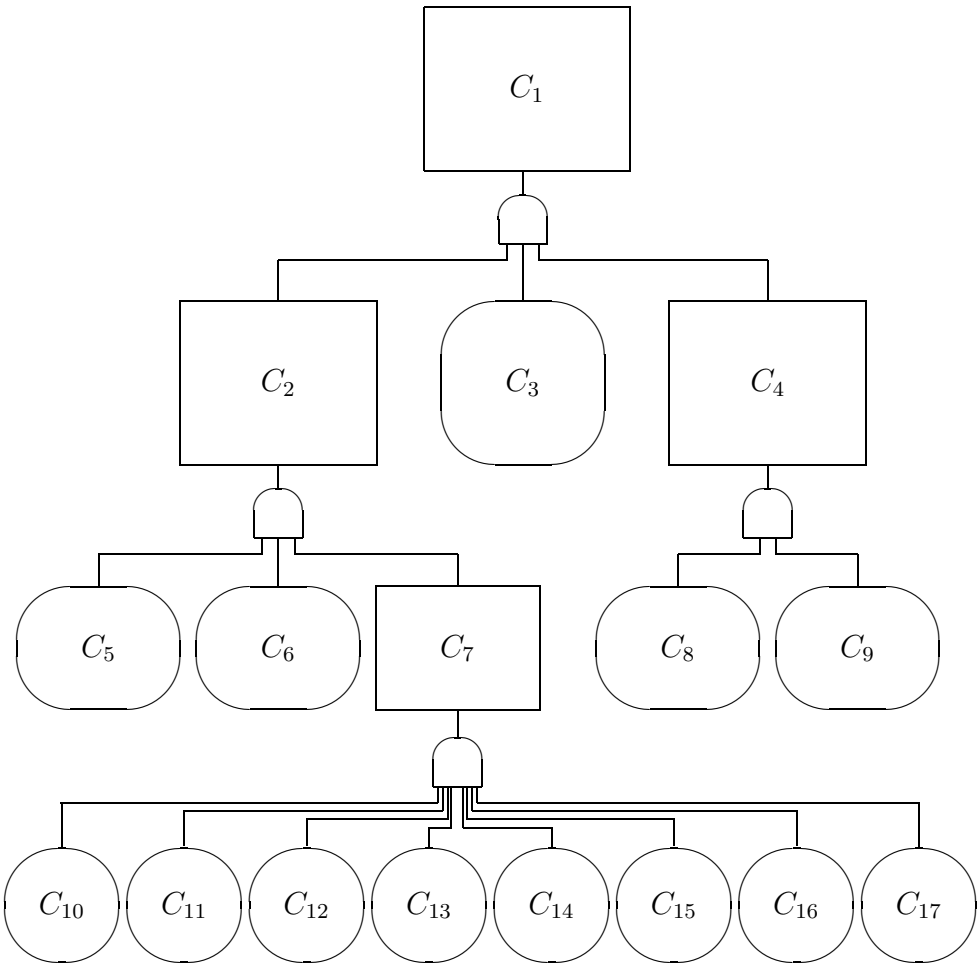

FIGURE 1. *Event Tree Diagram for Missile Reliability*

of detail should be included in the summary of the system and subsequent analysis is the sources of data available. For example, "component" C_3 is actually comprised of a collection of "subcomponents", which might have been alternately represented as a "subsystem" with "components", if data had been available on the subcomponents.

2.2. **Data Description.** In estimating the reliability of a complex system, it is common to have test data and prior engineering judgment available at the system, the subsystem, and the component levels. Methodology for combining these various sources of information in a consistent fashion has proven problematic (Bier, 1994), and the goal of this paper is to

describe an IIT approach that resolves this difficulty. For simplicity, we restrict discussion to systems in which components or subsystems are classified as either functional or not (pass/fail).

Four sources of data are considered here. The first is full-system data consisting of pass/fail results from 1249 flight tests. The second is data collected from component or subsystem tests, also assumed here to take on only pass/fail values. Several collections of observations are available, ranging from 70 to 100 observations for some, but not all, of the components and subsystems. Some of the data have covariate information, including storage location and the age of the system at test, which is not used in this analysis.

Third, engineering judgment regarding the probability that a specific component or subsystem fails may be available. It represents substantial subjective knowledge about the working of the system and its intended design that helps to bound the reliability.

A fourth, less precise source of information is engineering judgment stating that a group of components in a given system or in related systems have similar failure probabilities. For example, an expert may assert that the reliability of the missile battery is "similar" to the reliability of a battery in a related missile system, or that reliabilities of types of motors are similar. Alternately, an expert may judge that all of the components of a given subsystem are equally likely to cause the failure of that subsystem. This says that the reliability of the components are thought to be similar, not that the failure mechanisms of the components are the same.

3. Bayesian Reliability Modeling

3.1. **Background.** The new methodological advances that are the focus of this paper primarily involve how Bayesian models can be adapted to give consistent sytsem down and component up representation of success/fail reliability models given different granularity of the modeling, and some insights into how meaningful and informative Bayesian priors can

be formed to incorporate multiple sources of indirect supplementary information. To provide context, it is useful to begin with a review of some related research in Bayesian system reliability.

As with all statistical models used in system reliability, the first issue is the specification of the system structure, which formalizes how components and subsystems are related to each other and to the entire system. Many of the common ways of making this specification are graphical and include reliability block diagrams, event trees, fault trees, and Bayesian networks. For more details on any of these approaches see Rausand and Høyland (2004). Each of these representations implies a particular structure for the way that data are incorporated to make inferences about the system as a whole—this structure provides a means of obtaining the *likelihood*.

Often the likelihood contains unknown parameters. In classical or frequentist statistics, the standard approach is to find the value of the parameters that maximizes the likelihood of the observed data. These estimated parameter values can then be used in conjunction with the fitted model to obtain predictions and for other inference. Bayesian methodology allows one to augment the data with expert engineering knowledge via the specification of an a priori distribution for the unknown parameters, called the *prior distribution*. The data are then used to update the distribution on the parameters to a *posterior distribution* that captures both the information from the prior and from the observed data. As with classical inference, the posterior distributions allow prediction and inference. As the amount of observed data increases, the results from Bayesian and classical inference tend to converge, since information added through the prior distribution is downweighted relative to the data's contribution.

Martz, Waller and Fickas (1988) and Martz and Waller (1990) studied similar system reliability problems with binomial data, and also faced the problem of integrating expert opinion at different levels. They assumed that they were supplied with beta distributions for the reliabilities of individual components and of the subsystems made up of those components. Beta distributions for the components are then used to derive "induced" distributions for

subsystems (for example, a series subsystem is modeled as a product of random variables with the components' beta distributions). This "induced" distribution is then combined with the subsystem's beta distribution, and the combination is itself approximated using a beta distribution, with parameters determined by the method of moments. This procedure is repeated until the full system level is reached. Graves and Hamada (2004) reanalyze these data sets using the same Bayesian methods as in the present paper, which do not require approximations. A shortcoming of the Martz et al. (1988) method is that the full system data do not inform estimates of single component reliability. Because of this and since full system tests suggested lower reliability than the component and subsystem tests would indicate, Graves and Hamada (2004) obtained substantially smaller reliability estimates for some components.

Many reliability models do not consider prior expert opinion and data at multiple system levels. Springer and Thompson (1966, 1969) and Tang, Tang and Moskowitz (1994, 1997) provide exact or approximated system reliability distributions obtained by propagating the component posteriors through the system structure. Thompson and Chang (1975), Chang and Thompson (1976), Lampkin and Winterbottom (1983), and Winterbottom (1984) use approximations for exponential lifetimes rather than binomial data. Others propose methods for evaluating or bounding moments of the system reliability posterior distribution (Cole (1975), Mastran (1976), Dostal and Iannuzzelli (1977), Mastran and Singpurwalla (1978), Barlow (1985), Natvig and Eide (1987), and Soman and Misra (1993)). These moments can also be used in the beta approximations employed by Martz, Waller and Fickas (1988) and Martz and Waller (1990). Soman and Misra (1993) proposed a distributional approximation based on a maximum entropy principle.

3.2. **The Binary System and Component Model.** The model developed here assumes that the status of each component, subsystem and full system is in just one of two states: pass or fail. This status of the tested unit is the simplest outcome of a reliability test.

The restriction to binary status implies that independent testing results in a binomial likelihood. The representation of the full system as connected subsystems and components

in series makes the status of the system and subsystem functionally dependent on the status of the components. As a result, the parameters of the binomial likelihood of subsystems and the full system are functions of the parameters of the components. By coding a success as a one and a failure as a zero, the outcome is modeled by a Bernoulli random variable Y with distribution governed by a single parameter, the probability p it takes value one (namely, a success). The probability of a failure, the only other possible outcome, is therefore, $1 - p$. The goal of the modeling is to use the available data to estimate the parameters p_i for each component, subsystem, or full system, C_i, with their associated uncertainties. Primary interest likely lies in obtaining an estimate of the probability that the overall system C_1 succeeds, namely p_1.

As discussed previously, combining data and prior information at different levels within a reliability diagram has often proven problematic, both from the perspectives of computational tractability and model consistency. Our solution to this dilemma is to simply re-express nonterminal (subsystem) node probabilities in terms of terminal (component) node probabilities using deterministic relations derived from the system reliability diagram. For example, from Figure 1, it is evident that the probability that subsystem C_7 functions, p_7, is equal to the product of the probabilities that each of the components p_{10} through p_{17} all function. Thus,

$$p_7 = \prod_{i=10}^{17} p_i$$

Similarly, the probability that subsystem C_4 functions is $p_7 = p_8 \times p_9$.

Note that variable substitutions based on the reliability diagram do not uniquely identify a joint distribution on the terminal node probabilities, in this case p_{10} through p_{17}. However, this approach does use the marginal distribution of the system reliability to provide a sensible, although not necessarily unique, solution for the joint distribution of all of the component and subsystem reliabilities. See Johnson et al (2003) for additional details.

3.3. **Prior Specification.** In many applications, engineering judgment can play a important role in assessing system reliability, particularly for large complex systems where data

collected on individual components is sparse. Judgment and expertise are always required to develop the system representation and determine which data are relevant to the analysis.

In this analysis, we explore the statistical modeling appropriate for two additional types of judgment:

(1) The expert provides precise information about the reliability of a single component by specifying a probability distribution for the failure probability of the component.
(2) The expert identifies a component as belonging to a group of components with similar reliability. This assumption does not require that the components are physically similar, only that their reliabilities are similar. For example, all the high reliability components might be judged as being similar. In this case, the expert provides a family of probability distributions for the failure probability.

3.3.1. *Precise Information.* Engineering judgment may be available from several experts, and the quality of information obtained from each expert may vary. In the model, we therefore assume that the prior density obtained from expert m concerning a specific value of p_i takes the form of a beta density, and we let the set of combinations of (i, m) for which engineering judgment is available be denoted by S_1. A beta density is a flexible class of probability distributions that assigns mass between 0 and 1, thus making it suitable to capture judgment about probabilities. See Gupta and Nadarajah (2004) for more details on the beta distribution and how the choice of parameters influences the mean and variance of the distribution.

More specifically, we assume that the prior information is modeled as:

$$\begin{aligned} &\frac{\Gamma(N_m+2)}{\Gamma(N_m\pi_{i,m}+1)\Gamma[N_m(1-\pi_{i,m})+1]}p_i^{N_m\pi_{i,m}}(1-p_i)^{N_m(1-\pi_{i,m})} \\ &\qquad\qquad \equiv \mathrm{B}(p_i\,;\, N_m\pi_{i,m}+1, N_m(1-\pi_{i,m})+1). \end{aligned} \tag{1}$$

Note that this is not the standard parameterization of the beta distribution, but is selected for the desirable feature that it has the mean of the distribution as one of the parameters,

$\pi_{i,m}$. To equate the more standard $\mathrm{B}(a, b)$ with mean $a/(a+b)$, set $a = N_m\pi_{i,m} + 1$ and $b = N_m(1 - \pi_{i,m}) + 1$. See Gupta and Nadarajah (2004) for details.

In equation (1), $\pi_{i,m}$ represents expert m's point estimate of p_i, and N_m represents the precision of expert m. The parameter N_m is unknown, and for concreteness, we assume that each expert precision parameter N_m is drawn from a gamma density with known parameters α_m and β_m, parameterized here as

$$\mathrm{G}(N_m\,;\,\alpha_m, \beta_m) = \frac{\beta_m^{\alpha_m}}{\Gamma(\alpha_m)} N_m^{\alpha_m - 1} \exp(-\beta_m N_m).$$

In summary, engineering judgment that a particular parameter value is approximately $\pi_{i,m}$ can be captured by a beta distribution. The precision of the engineering judgment is unknown a priori; the parameter capturing the precision is N_m, and is estimated from data, but with an informative prior. The prior distribution on N_m is a gamma distribution, which allows additional flexibility to be added to adjust the variability of the estimate to one that appropriately reflects system knowledge. By estimating N_m, some assessment of the value added by a given expert's judgement can be made. An effect of estimating N_m is that the expert's information will be downweighted if it is inconsistent with data. Note that engineering judgment has the form of a binomial likelihood with a maximum at $\pi_{i,m}$. This convention eliminates the possibility that the joint density specified on all model parameters is improper, and also implicitly handles the aggregation problem identified by Bier (1994) by simply treating engineering judgment as a form of "data."

Simple examples of the differences between fixed and random N are illustrated in Figure 2. In each case, the data are x successes in $n = 50$ binomial trials with unknown success probability p. The prior distribution for p is $B(N\pi, N(1-\pi))$ for $\pi = 0.5$, and we compare the case where N is fixed at 50 with the case where N has a gamma prior distribution with parameters $\alpha = 5$ and $\beta = 1/10$ (so that $E(N) = 50$). In each plot, the solid curve is a posterior density estimate for p in the fixed N case, while the dotted-dashed curve is a density estimate for p in the random N case. The left plot is for the case of $x = 25$ successes, which is exactly consistent with the prior mean of $\pi = 0.5$. In this case, the data are not

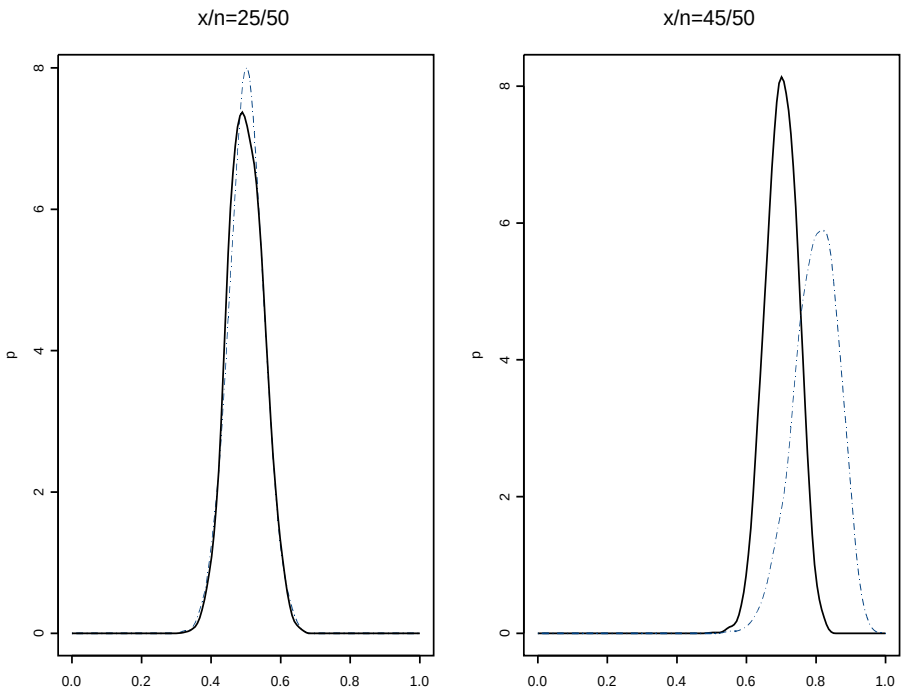

FIGURE 2. Comparison of fixed and random prior sample sizes. Shown are posterior densities of p with a $B(N\pi, N(1-\pi))$ prior with $\pi = 0.5$ based on $x = 25$ (left) or $x = 45$ (right) successes in $n = 50$ trials, for the cases of fixed $N = 50$ (solid curves) or random N with a prior distribution with mean 50 and SD 22 (dotted-dashed curves).

very informative about N so that its posterior is close to its prior; the expert could be very reliable (large N) or unreliable and fortunate to have guessed the right value (small N). The posterior mean of N is 52, slightly larger than its prior mean of 50, and its posterior standard deviation (SD) is about the same as its prior SD of 22.4. The posterior mean of p is 0.5, and its posterior SD is essentially 0.05 in either the fixed or random N case. Therefore no harm is done by allowing N to be random in the case of data that agree with the prior. However, if the data and the prior disagree, as in the right plot of Figure 2 with $x = 45$, the random N analysis gives different results. For fixed $N = 50$, the inference is like 70 successes in 100 trials so $E(p) = 0.7$ and $SD(p) = 0.046$. When N is random, its posterior mean of N is 18 with SD 11, so that the prior for p is deemphasized and we get $E(p) = 0.8$ and $SD(p) = 0.063$; the mean is closer to the high reliability implied by the data, while the standard deviation is larger because the prior is contributing fewer data points. The fixed N

analysis is misleading because the accuracy of the prior has been overestimated, as is shown by its disagreement with the data, and the random N analysis behaves accordingly, treating the prior as useful information but not as accurate as initially thought.

3.3.2. *Component Groupings.* When prior information regarding component success probabilities is unknown, but groupings of "similar" components are available, equation (1) is augmented in the model by assuming that $\pi_{i,m}$ is replaced by $\rho_{m,g}$, where $\rho_{m,g}$ represents the common, but unknown, success probability assigned by expert m to components in the broader group g. The form of the prior on model parameters from such information takes the form

$$\prod_{(i,m)\in S_2} \mathrm{B}(p_i\,;\, K_m\rho_{m,g}+1, K_m(1-\rho_{m,g})+1).$$

Here, S_2 denotes the combinations of (i,m) for which such grouping information is available.

As in equation (1), the parameter K_m is assumed to be drawn *a priori* from a gamma density having parameters ζ_m and η_m. The prior success parameter $\rho_{m,g}$ for a particular group is assumed to be from a beta density with known parameters $\delta_{g,m}$ and $\epsilon_{g,m}$, respectively. Here $\rho_{m,g}$ can be interpreted as a common mean estimate of the individual p_i's for elements within that group.

This approach allows for knowledge from different components to be leveraged across components deemed to be similar, to increase predictive power and exploit common features in the system.

3.3.3. *Hierarchical Model to Address Data Granularity.* In this analysis, a hierarchical prior is used on the components/terminal nodes. This prior is used not to capture expert judgment, but to render estimates of the overall system reliability insensitive to the level of detail included in the system event diagram.

As an illustration of this point, consider a simple system comprised of three components connected in series (similar to a system, C_1, comprised only of C_2, C_3 and C_4, with no other components) and suppose that a single binomial observation with four successes and one failure is observed at the system level. Without a hierarchical specification on the

component probabilities and under the model assumptions stated above with uniform priors, the likelihood of the system reliability would be proportional to

$$(p_2p_3p_4)^4(1-p_2p_3p_4)$$

where the system reliability, p_1, is assumed equal to $p_2p_3p_4$. This is obtained using the fact that since four of the trials resulted in successes, implying that components C_2, C_3 and C_4 all worked, which occurs with probability $p_2p_3p_4$. The trial which resulted in a failure had probability, $1 - P(Success) = 1 - p_2p_3p_4$.

If we assume a uniform distribution for the prior of each of p_2 to p_4, the posterior mean of p_1 in this model is 0.507. When the system is not decomposed into subsystems and a uniform prior is assumed on p_1, the posterior mean on p_1 (with a uniform prior) is 0.714. Furthermore, under such naive model specifications, the bias attributable to adding components to the event tree becomes more severe as the number of components in the system increases.

Suppose instead that the reliabilities of components C_2, C_3, and C_4 are not assumed to have independent uniform distributions. Instead, each component's success probability is drawn from a beta density (as given in equation (2) with parameters $J\gamma$ and $J(1-\gamma)$, where the parameter J is assumed drawn from a gamma density function with parameters $\boldsymbol{\tau}$ and $\boldsymbol{\phi}$. The parameter γ is assumed to be drawn from a beta density with parameters $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$.

$$\frac{\Gamma(J)}{\Gamma(J\gamma)\Gamma[J(1-\gamma)]}p_i^{J\gamma}(1-p_i)^{J(1-\gamma)} \equiv \mathrm{B}(p_i\,;\,J\gamma, J(1-\gamma)). \tag{2}$$

Using this hierarchical prior specification on p_2 to p_4 with $\psi = \omega = 0.5$ results in a posterior mean of 0.718 for p_1, while the same specification with $\psi = \omega = 1.0$ results in a posterior mean of 0.687. Both estimates are largely insensitive to the number of components specified for the system. This is an important feature of this model, as the availability of

data for different components and subsystems in the system should not play an important factor in influencing overall reliability estimate of the system.

3.3.4. *Joint Distribution.* To obtain a posterior distribution for the Bayesian analysis, the likelihood which contains the information captured in the data must be combined with the expert judgement summarized in the prior distributions. The posterior is proportional to the product of the likelihood times the prior distributions. For more details, see Martz and Waller (1982). Combining the information from the likelihood and priors leads to a joint posterior distribution on the model parameters proportional to

$$
\begin{aligned}
&[\boldsymbol{p}, \mathbf{N}, \boldsymbol{\rho}, \mathbf{K}, \gamma, J \,|\, \mathbf{x}, \mathbf{n}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, \boldsymbol{\tau}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\omega}] \propto \\
&\prod_{i \in S_0} p_i^{x_i} (1-p_i)^{n_i} \\
&\times \prod_{(i,m) \in S_1} \mathrm{B}(p_i \,;\, N_m \pi_{i,m} + 1, N_m(1-\pi_{i,m}) + 1) \\
&\times \prod_{m : \exists (i,m) \in S_1} \mathrm{G}(N_m \,;\, \alpha_m, \beta_m) \\
&\times \prod_{(i,m) \in S_2} \mathrm{B}(p_i \,;\, K_m \rho_{m,g} + 1, K_m(1-\rho_{m,g}) + 1) \\
&\times \prod_{m : \exists (i,m) \in S_2} \mathrm{B}(\rho_{m,g} \,;\, \delta_m, \epsilon_m) \times \prod_{m : \exists (i,m) \in S_2} \mathrm{G}(K_m \,;\, \zeta_m, \eta_m) \\
&\times \prod_{i \in S_0} \mathrm{B}(p_i \,;\, J\gamma, J(1-\gamma)) \\
&\times \mathrm{B}(\gamma \,;\, \boldsymbol{\psi}, \boldsymbol{\omega}) \mathrm{G}(J \,;\, \boldsymbol{\tau}, \boldsymbol{\phi}).
\end{aligned} \tag{3}
$$

In equation (3), we have an expression for the joint posterior distribution of the parameters of primary interest $(\boldsymbol{p}, \mathbf{N}, \boldsymbol{\rho}, \mathbf{K}, \gamma, \mathbf{J})$, which are shown to the left of the vertical line on the left hand side of the expression. The vertical line indicates that we are conditioning on the other parameters of the model $(\mathbf{x}, \mathbf{n}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, \boldsymbol{\tau}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\omega})$ to obtain an expression

that is proportional ($\propto$) to the joint posterior distribution. S_0 denotes the set of terminal nodes/components.

In this expression, values of nonterminal node probabilities (such as those for elements C_1, C_2, C_4 and C_7) are assumed to be expressed in terms of the appropriate functions of terminal node probabilities, as defined from the system event diagram. The first line of equation (3) corresponds to the contribution from observed data for each available component from the likelihood function. The second and third lines correspond to the contribution of expert opinion for individual components through the component priors, while lines 4 and 5 summarize the information from experts on groups of components.

An examination of the contributions to the joint posterior distribution arising from the prior information reveal obvious similarities, but there are also important distinctions between these parameterizations. For example, the value of N_m represents the precision of the expert's opinion, while K_m describes the similarity of item reliabilities within a grouping.

3.4. **Implementation: Markov Chain Monte Carlo.** The joint posterior distribution in equation (3) does not have a familiar distributional form, so it is not immediately obvious how to do inference and estimation. Recent advances in Bayesian computing allow one to obtain a random sample from the joint posterior distribution. Once one has a random sample, then inference can be made on any of the quality characteristics of interest. Gibbs sampling (Casella and George 1992) is one method to draw random samples from the joint posterior distribution. Another method is the Metropolis-Hastings algorithm (Chib and Greenberg 1995).

One issue with Gibbs sampling is whether the draws are approximately a random sample from the posterior distribution; this is referred to as the convergence of the Gibbs sampler. To mitigate the impact of initial values chosen for the parameters, a burn-in is typically performed in which the Gibbs sampler is run a number of times and the draws thus obtained are discarded. To reduce dependence between draws, the draws can be thinned by retaining every kth draw. See Raftery and Lewis (1996) for more discussion of diagnostics for convergence.

The joint distribution on model parameters specified in Section 3.3.4 does not lend itself to analytical evaluation of the system or component reliabilities. However, a componentwise Metropolis-Hastings algorithm can be implemented in a relatively straightforward way. In our version of such a scheme, we use a random-walk Metropolis-Hastings algorithm with Gaussian proposal densities specified on the logistic scale for the terminal node probabilities, as well as for ϱ_0 and $\rho_{m,g}$. Precision parameters are similarly updated using a random-walk Metropolis-Hastings scheme with Gaussian increments specified on the logarithmic scale. The resulting Metropolis-Hastings algorithms are implemented using YADAS (Graves, 2001 and 2003).

3.5. **Posterior and Predictive Distribution.** Using engineering judgment, reliability classes are formed as follows. System elements C_2-C_4 are assigned to Group 1, elements C_5-C_9 to Group 2, and C_{10}-C_{17} to Group 3. These groupings are handled as in Section 3.3.2. Note that there is great flexibility about the creation of groups, as some of them include both components and subsystems; also, the groupings need not be mutually exclusive. Beta distributions with common, fixed group means (π_1-π_3) and a single, common precision parameter ($N_{1,2,3}$) are assumed for each of Groups 1-3. A common precision parameter is incorporated for each group, since a single expert provided all this information. The individual components in this system are C_3,C_5,C_6 and C_8-C_{17}. As described in Section 3.3.3, a hierarchical prior with unknown mean and precision parameter (γ and J, respectively) is assumed for components in Group 4. Also, gamma priors with parameters (5,1) are assumed for both precision parameters ($N_{1,2,3}$ and J), and a noninformative prior ($\psi = \omega = 0.5$) is assumed for γ.

Applying the model discussed in Section 3.3.4, we obtained the posterior distributions on the component reliabilities for each of the components and the expert precision parameters. The system reliability posterior distributions with the system data included and system data excluded are plotted in Figure 3. It is clear that when the flight test data are included a much more precise estimate of system reliability can be obtained, as noted by the much narrower posterior distribution. However, it should also be noted that the posterior distribution

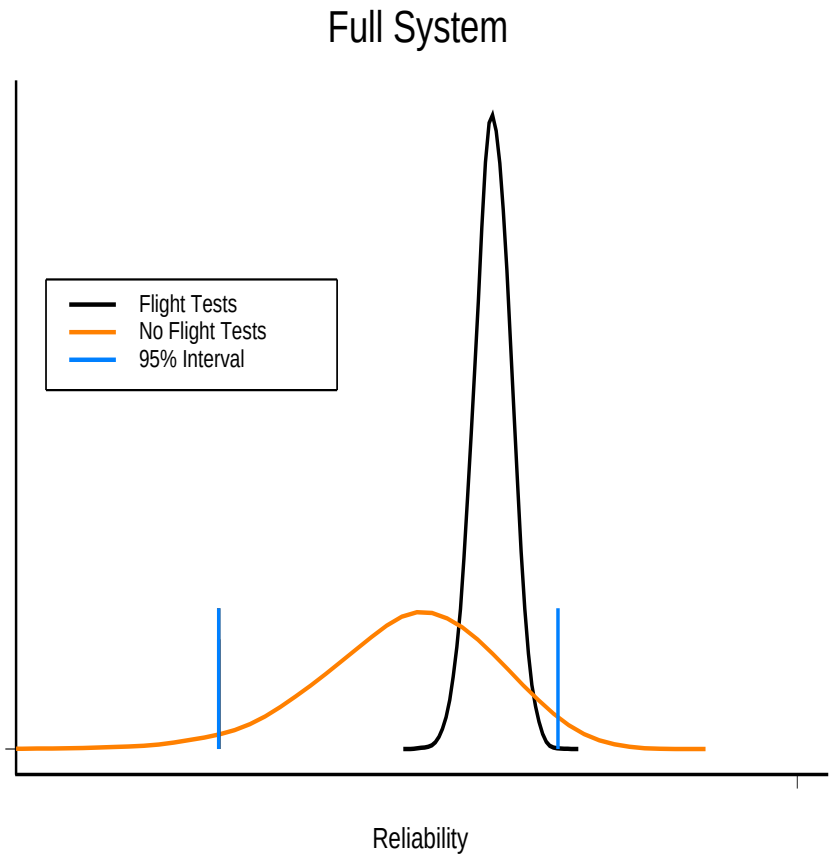

FIGURE 3. Posterior distributions for the reliability of the system represented in Figure 1. The solid black lines are based on the model that included the full-system flight tests; the dashed lines show results without full-system level data. The 95% interval is based on results without full-system level data

obtained without the benefit of the flight data, and only based on the indirect component and subsystem data, is able to appropriately estimate a sensible range of system reliability values.

Posterior distributions for all components are given in Figure 4. In the cases where the posterior distributions are very similar, such as C_3, these correspond to cases where component level data are available. When the marginal posterior distribution based on component-level data only is substantially more disperse than the posterior based on results using the flight data, such as C_1, this corresponds to situations where no components level data is available. In these cases, the flight data is highly beneficial since the entire system working coupled

with the series form of the system provides information that these components also worked. We note the agreement between the two posterior distributions (full system tests included vs full system tests excluded). In every case, the 95% highest posterior density (HPD) region includes essentially the entire distribution with full system information included. This validates that the full system data and the combined component data are summarizing system reliability similarly, but just with different precision. In addition, it is worth noting that this approach can provide a system estimate of reliability even in the absence of full system data, such as is encountered for the nuclear stockpile. Note that for this case, since there is a great deal of flight test data, the data dominates the prior in the first calculation.

Also of interest is the posterior distribution for the expert precision parameter $N_{1,2,3}$. The posterior mean for this distribution is 12.2. This suggests that the expert's opinion is worth approximately 12 full system tests. Given the prior mean of 5, we conclude that the expert is reasonably well calibrated with the system structure and data. This is also a useful measure to determine the added benefit of including expert opinion in the analysis.

3.6. **Diagnostics.** Two concerns commonly encountered in modeling system-level reliabilities using series diagrams like that depicted in Figure 1 involve the extent to which different components function independently and whether system (or subsystem) reliability decreases when components are assembled. A simple cross-validation diagnostic useful for assessing the importance of these influences can be constructed by iteratively omitting data collected at each node from the estimation procedure, and then examining the predictive distribution for the omitted datum.

Such a procedure is applied to data obtained for this missile system and resulted in an estimate of 0.83 for the predictive probability of observing fewer successes at the system-level than are actually observed. It therefore seems that there is little evidence to support the notion that the reliability of the system is degraded as components are assembled and required to operate as a unit.

There is, however, some indication of model lack-of-fit at the component level. For components 10 and 17, the predictive distribution for observing fewer successes than are obtained

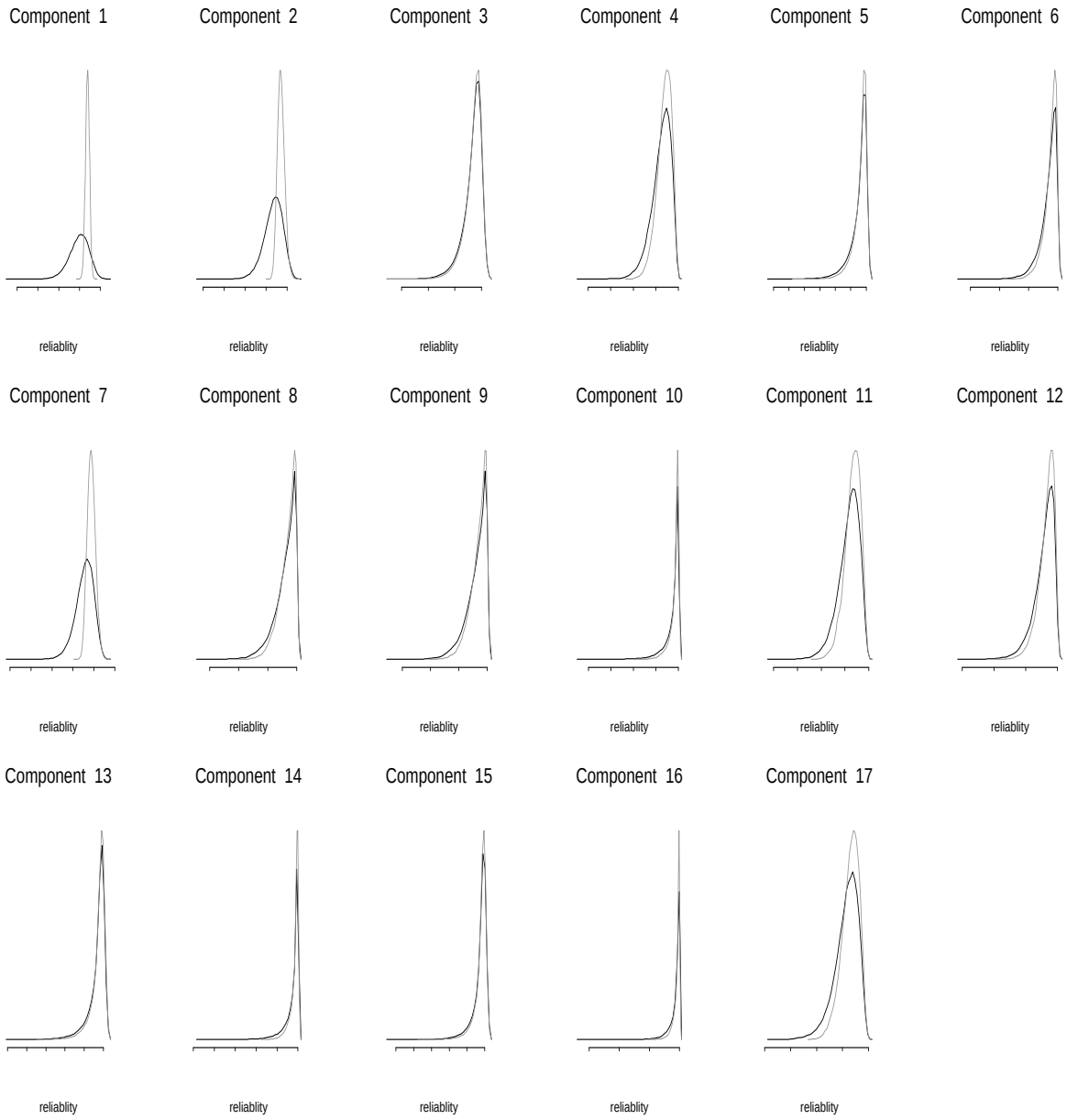

FIGURE 4. Posterior distributions for the reliability of the system represented in Figure 1. In each pair of plots, the more peaked curve represents the marginal posterior density based on all test data, while the more dispersed curves represent the marginal posterior density using only component-level data (i.e., excluding system-level tests).

at these nodes is approximately 3.5%. The same number of failures are observed at each of these components, and these two components had the highest failure rate of any components in the system. Model lack-of-fit in this instance might thus be attributed to the fact that the hierarchical mean estimated for the terminal nodes, γ, increased substantially when the

datum for either of these nodes is omitted, resulting in an overly optimistic estimate of this probability. Possible remedies for such model inadequacy would be to stochastically decrease the prior assigned to the value of J, or to introduce a separate hierarchical group for these nodes. In this case, neither remedy appeared to substantially affect estimates of system reliability in subsequent sensitivity analyses.

4. Conclusions

In this paper, a Bayesian approach to combining component, subsystem and system data with expert judgement is presented. It allows for the flexible combination of multiple sources of data with different weights for the various sources of information. In addition to being computationally manageable with the use of freely available software, the approach reduces the final estimates dependence on which components have data available. This robustness to the structure of available data is appealing, since frequently the availability of data at various components is a function of cost or convenience.

The approach also allows for comparison of results between the complete analysis and an analysis when full system data is not incorporated. The similarity of the two sets of results for this system validate that the system has been appropriately modeled, and provides insights into the usefulness of this approach when full system data may not be available.

References

[1] Barlow, R. E. (1985), "Combining Component and System Information in System Reliability Calculation," in *Probabilistic Methods in the Mechanics of Solids and Structures*, Eds. S. Eggwertz and N.L. Lind, Springer-Verlag, Berlin, 375-383.

[2] Bier, V. M. (1994), "On the Concept of Perfect Aggregation in Bayesian Estimation," *Reliability Engineering and System Safety*, **46**, 271-281.

[3] Casella, G. and George, E. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, **46:3**, 167-174.

[4] Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, **49:4**, 327-335.

[5] Cole, P. V. Z. (1975), "A Bayesian Reliability Assessment of Complex Systems for Binomial Sampling," *IEEE Transactions on Reliability*, **R-24**, 114-117.

[6] Dostal, R. G. and Iannuzzelli, L. M. (1977), "Confidence Limits for System Reliability When Testing Takes Place at the Component Level," in *The Theory and Applications of Reliability* (Vol. 2), Academic Press, New York, 531-552.

[7] Gilks, W. R. Thomas, A. and Spiegelhalter, D. J. (1994), "A Language and Program for Complex Bayesian Modelling," *The Statistician*, **43**, 169-178.

[8] Graves, T. L. (2001), "YADAS: An Object-Oriented Framework for Data Analysis Using Markov Chain Monte Carlo," Los Alamos National Laboratory Technical Report, LA-UR-01-4804.

[9] Graves, T. L. (2001), "An Introduction to YADAS," yadas.lanl.gov.

[10] Graves, T. L. and Hamada, M. S. (2004), "Combining Multi-Level Data to Assess System Reliability and Allocate Resources Optimally: Bayesian Methods and Computation." Los Alamos National Laboratory Technical Report, LA-UR-04-7157.

[11] Gupta, A. K. and Nadarajah, S. (2004), *Handbook of Beta Distribution and Its Applications*, New York: Marcel Dekker, Inc.

[12] Johnson, V. E., Graves, T. L., Hamada, M. and Reese, C. S. (2003), "A Hierarchical Model for Estimating the Reliability of Complex Systems (with Discussion)," *Bayesian Statistics 7*, Oxford University Press.

[13] Martz, H. F. and Waller, R. A. (1982), *Bayesian Reliability Analysis*, New York: Wiley.

[14] Martz, H. F. and Waller, R. A. (1990), "Bayesian Reliability Analysis of Complex Series/Parallel Systems of Binomial Subsystems and Components," *Technometrics*, **32**, 407-416.

[15] Martz, H. F. Waller, R. A., and Fickas, E. T. (1988), "Bayesian Reliability Analysis of Series Systems of Binomial Subsystems and Components," *Technometrics*, **30**, 143-154.

[16] Natvig, B. and Eide, H. (1987), "Bayesian Estimation of System Reliability," *Scandinavian Journal of Statistics*, **14**, 319-327.

[17] Raftery, A. E. and Lewis, S. M. (1992), "How many iterations in the Gibbs sampler?" in *Bayesian Statistics 4*, eds. J. M. Bernardo et al., Oxford University Press, 763-773.

[18] Rausand, M. and Høyland, A. (2004), *System Reliability Theory: Models, Statistical Methods, and Applications*, New York: Wiley.

[19] Roberts, N. H. (1987), *Fault Tree Handbook*, United States Government Printing.

[20] Soman, K. P. and Misra, K. B. (1993), "On Bayesian Estimation of System Reliability," *Microelectronics and Reliability*, **33**, 1455-1459.

[21] Springer, M. D. and Thompson, W. E. (1966), "Bayesian Confidence Limits for the Product of N Binomial Parameters," *Biometrika*, **53**, 611-613.

[22] Springer, M. D. and Thompson, W. E. (1969), "Bayesian Confidence Limits for System Reliability," in *Proceedings of the 1969 Annual Reliability and Maintainability Symposium*, Institute of Electrical and Electronics Engineers, New York, 515-523.

[23] Tang, J., Tang, K., and Moskowitz, H. (1994), "Bayes Credibility Intervals for Reliability of Series Systems with Very Reliable Components," *IEEE Transactions on Reliability*, **43**, 132-137.

[24] Tang, J., Tang, K., and Moskowitz, H. (1997), "Exact Bayesian Estimation of System Reliability from Component Test Data," *Naval Research Logistics*, **44**, 127-146.

Statistical Science Group, Los Alamos National Laboratory